DOCUMENT RESUME

ED 058 612 EA 003 912

AUTHOR	Levine, R. A.; Williams, A. P., Jr.
TITLE	Making Evaluation Effective: A Guide.
INSTITUTION	Carnegie-Mellon Univ., Pittsburgh, Pa.; Rand Corp., Santa Monica, Calif.
SPONS AGENCY	Department of Health, Education, and Welfare, Washington, D.C.
REPORT NO	R-788-HEW-CMU
PUB DATE	May 71
NOTE	48p.
AVAILABLE FROM	Communications Department, Rand, 1700 Main Street, Santa Monica, California 90406 ($2.00)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Decision Making; *Educational Planning; *Evaluation; Evaluation Criteria; *Evaluation Methods; Evaluation Needs; *Evaluation Techniques; Federal Programs; Government Role; Guidelines

ABSTRACT

This manual was prepared to aid in building an effective evaluation structure for government programs and projects throughout the HEW agencies. It is intended to assist evaluators in the development of a strategy -- when to evaluate, who should evaluate, and how to supervise the evaluation; how to allocate resources among different types of evaluation; and how to assess evaluation. The presentation avoids questions of methodology and concentrates on the decisionmaking context for evaluation. Suggested documentation includes an annual evaluation plan and statement of strategic objectives to be prepared by each evaluation office, a statement of the rationale of each evaluation to be prepared by the initiating office, and a critical program summary to be prepared by the chief evaluation officers. (Author)

This report was sponsored by the Department of Health, Education and Welfare under Subcontract HEW5-70-155 between Carnegie-Mellon University and The Rand Corporation. Views or conclusions contained in this study should not be interpreted as representing the official opinion or policy of HEW, Rand or the Carnegie-Mellon University.

ED 058612

OE - DCM
EA

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

R-788-HEW/CMU

May 1971

MAKING EVALUATION EFFECTIVE: A GUIDE

R. A. Levine and A. P. Williams, Jr.

A Report prepared for

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

EA 003 912

Rand
SANTA MONICA, CA. 90406

Rand maintains a number of special subject bibliographies containing abstracts of Rand publications in fields of wide current interest. The following bibliographies are available upon request:

Africa • Arms Control • Civil Defense • Combinatorics
Communication Satellites • Communication Systems • Communist China
Computing Technology • Decisionmaking • Delphi • East-West Trade
Education • Foreign Aid • Foreign Policy Issues • Game Theory
Health-related Research • Latin America • Linguistics • Maintenance
Mathematical Modeling of Physiological Processes • Middle East
Policy Sciences • Pollution • Program Budgeting
SIMSCRIPT and Its Applications • Southeast Asia • Systems Analysis
Television • Transportation • Urban Problems • USSR/East Europe
Water Resources • Weapon Systems Acquisition
Weather Forecasting and Control

To obtain copies of these bibliographies, and to receive information on how to obtain copies of individual publications, write to: Communications Department, Rand, 1700 Main Street, Santa Monica, California 90406.

Published by The Rand Corporation

$2.00

R-788-HEW/CMU Making Evaluation Effective: A Guide. R. A. Levine, A. P. Williams, Jr. May 1971.

Text of an evaluation manual prepared for the Deputy Assistant Secretary for Evaluation and Monitoring of the Department of Health, Education, and Welfare to aid in building an effective evaluation structure for government programs and projects throughout the agencies of HEW. This guide is intended to assist evaluators in the development of a strategy--in determining when to evaluate, who should evaluate, and how to supervise the evaluation; how to allocate efforts among different types of evaluation; and how to assess evaluation. As per the HEW request, the guide avoids questions of methodology and concentrates instead on the decisionmaking context for evaluation. Suggested documentation includes an annual evaluation plan and statement of strategic objectives to be prepared by each evaluation office, a statement of rationale of each evaluation to be prepared by the initiating office, and critical program summary to be prepared by chief evaluation officers.

48 pp. (KB)

*****

GOVERNMENT 2

guide for making evaluation effective within agencies of HEW

DECISIONMAKING

same

EVALUATION METHODS

same

EDUCATION PLANNING

same

3

CMU

Adm/Econ

EDUCATION

FILMED FROM BEST AVAILABLE COPY

PREFACE

This Report contains the text of an Evaluation Manual prepared in response to the need of the Deputy Assistant Secretary for Evaluation and Monitoring of the Department of Health, Education and Welfare for such a manual to aid in building an effective evaluation structure throughout the agencies of HEW.

The HEW request was for a guide not to the techniques of evaluation, but to the purposes of evaluation and how evaluation programs could meet these purposes. For this reason, the manual tries insofar as is possible to stay away from questions of methodology, concentrating instead on the decisionmaking context for evaluation.

Nor does the manual attempt to lay out rules for evaluators which, if followed, would get them to do evaluations "right." Conceptually, it might have been possible to provide decision rules for likely contingencies at a level of detail such that evaluators at all levels of HEW would have carried out relevant and competent evaluations simply by following those rules. In reality, however, this is quite impossible; perhaps it never will be possible. Rather than setting forth such rules, then, the text is written so that if an evaluator follows it he will provide explicit answers--for the record--to the questions that should be asked by those who plan and carry out evaluations. Such answers will not in themselves produce good evaluations, although they should help by directing the evaluator to important considerations he might otherwise neglect. Even more important, however, such on-the-record answers to the relevant questions will enable administrators and overseers of evaluation to determine whether the kind of evaluation they want is being done, and to institute the necessary corrections if they feel that it is not. A manual that provides this sort of basis for further action is less ambitious than a manual that in itself directs what should be done; the latter, however, would be neither usable nor used.

The manual has been prepared under a subcontract with Carnegie-Mellon University.

ACKNOWLEDGMENTS

The authors are indebted to Kathleen A. Archibald, H. A. Averch, G. D. Brewer, Sue A. Haggart, O. Hoeffding, Marjorie L. Rapp, and Walter Williams for their help in preparing this report and their criticisms of it. The authors are entirely responsible for both its content and format.

CONTENTS

PREFACE iii

ACKNOWLEDGMENTS v

Section
I. INTRODUCTION 1

II. TYPOLOGY OF EVALUATION AND RELATED ACTIVITIES 3
A. For What Purpose? 3
B. What Is Being Evaluated? 8
C. For Whom? 8

III. ANNUAL EVALUATION PLANNING 10
A. Statement of Strategic Objective 11
B. Evaluation Briefs 12
C. Tabular Summaries 13

IV. THE RATIONALE FOR A SPECIFIC EVALUATION 18
A. Where in the Typology of Part II Does the Evalation Fit? 18
B. What Decision or Decisions May Be Affected by the Evaluation Report? 19
C. How Might Possible Alternative Results of the Evaluation Affect the Decision? 21
D. Who Will Make the Final Decision? 21
E. How Will the Availability of Evaluation Results Match the Decision Timetable? 24
F. Will the Evaluation Results be Read and Understood? 27
G. Are the Evaluation Results Likely to be Misinterpreted? 29
H. If an Evaluation-based Decision Is Made, Can It Be Implemented? 31
I. Why Do Decisionmakers Want the Evaluation Results? . 31
J. Who Has the Managerial and Technical Competence and Time to Do the Evaluation? 32
K. How Credible is Evaluation? 34
L. How Does the Cost of the Evaluation Compare with the Program Dollars to be Affected by the Decision?..... 35

V. THE CRITICAL PROGRAM SUMMARY 37
A. What Is the Program in Question, and to What Problems is it Addressed? 37
B. What Do We Know about the Program? 37
C. What Do we Think about the Program? 38
D. What Action Do We Recommend in Regard to the Program? 38

E. What Additional Knowledge Would We Like to Have about the Program? 38
F. What Should Be Done to Obtain this Information? 39

Appendix
A. Document Summary 40

I. INTRODUCTION

Evaluation, in its most general meaning, is a process of measurement designed to estimate worth. The definition does not imply any answer as to why worth is being estimated. As viewed from a government evaluation office, however, the "why" question is central, and a narrower meaning must be given to the word "evaluation." Evaluation of government programs or projects is a process of assessment designed to provide information about past and present operations and effectiveness, in order to assist in making decisions about the future.

Within the Federal government, evaluation offices often have additional responsibility for activities that, while related to evaluation, are distinct from it. Within the Department of Health, Education and Welfare, evaluation offices include among their activities the development of information or data systems and some monitoring or compliance control activities. Since these activities often compete with evaluation for the same funds, they will be considered explicitly in this Guide.

Because the importance of evaluation to the government is in the assistance it gives to decisionmaking, it must be planned and designed with the decision function in mind. In addition to being technically competent, government program evaluation must thus be relevant to decisionmaking and decisionmakers, and it must be credible. Credibility is based in part on technical competence and in part on the evaluator's reputation for independent appraisal.

This is not a Guide on how to evaluate. It is, rather, a Guide to assist evaluators in the development of a strategy: in determining when to evaluate, who should evaluate and supervise the evaluation, how to allocate efforts among different types of evaluation; and in assessing evaluation. The Guide has four objectives:

1. To aid in the planning of evaluation and related activities so that the limited resources available to evaluation offices are put where they are most needed.

2. To aid in carrying out evaluations in ways that are most likely to help in the making of decisions.

3. To aid in assessing and improving the total evaluation effort.

4. To aid in using evaluation as a part of the policymaking process.

Section II below presents a typology of evaluations--a mode of classification that can be used as an aid to the evaluation planning and execution processes with which the Guide is primarily concerned. Section III covers the planning of evaluation; Section IV, the preparation for specific evaluations; and Section V, the preparation of Critical Program Summaries, which brings together the evaluator's views on the programs he is responsible for. An appendix summarizes the documentation called for by this Guide.

II. TYPOLOGY OF EVALUATION AND RELATED ACTIVITIES

Evaluations can be classified in a number of ways: by what is being evaluated, by who conducts the evaluation, by the decision that is supposed to be affected by the evaluation, by the methodology used. The appropriate classification depends, of course, on the purposes for which it is to be used. The classification here provides a beginning; undoubtedly its use will indicate needed changes.

This typology is designed to be used in planning an evaluation program, to provide summary descriptions for planning the allocation of available evaluation funds and efforts. It should also be used in the preparation of a specific evaluation, as a means of shorthand notation on the functions of the project. In each case, it is necessary to indicate the purpose of the evaluation, what is being evaluated, and for what decisionmaking official the evaluation is being made. These three dimensions of the typology are discussed below.

A. FOR WHAT PURPOSE?

Evaluation and monitoring funds within the Department of Health, Education and Welfare are expended in the service of three broad objectives: evaluation proper, intended to aid in making decisions about a program or a treatment within a program; monitoring, more accurately described as compliance control; and building of capabilities for further evaluation. The distinctions among these categories, and among their subcategories--particularly the two major subcategories of evaluation proper--depend in large measure on <u>why</u> the evaluation project is being planned or set up. The key is the decision it is hoped the project will affect: if the project is designed to have impact on a decision concerning program size or program strategy, it belongs in one of the subcategories of evaluation proper; if it is to check adherence to regulations, it is compliance control; if its aim is not to contribute directly to program functioning, but rather to improve future evaluation efforts, it should be classed as capability building.

1. Evaluation

a. Evaluation to Affect Resource Allocation. This is the classical kind of evaluation designed to assess the worth or effectiveness of an ongoing program or project with a view to helping determine the quantity of funds or other resources it should receive. In the limit, the choice may be between no resources and some--a "go/no-go" decision to continue or cut off a project or program. More ordinarily, however, the decision based on such an evaluation is one of resource allocation at the margin--putting more into the "best" programs, holding back or cutting back on the less worthy.

Evaluations in this category vary with respect to the basis of comparison used in assessing relative worth: the impact or effectiveness of a program may be compared with unmet needs, with specified performance objectives, with past program performance, with the performance of equivalent programs, and/or with its own costs. Comparison with equivalent programs on the basis of program costs and of results measurable in the same units (e.g., percent of students in an in-school program whose reading scores have been raised by one grade level as compared with the same percent for program carried on outside the regular school system) is sometimes termed cost-effectiveness evaluation. The implicit decision to which such an evaluation is addressed is that of allocation among the programs whose costs and effectiveness are being compared. If program results can be expressed in dollar terms comparable not only to other program results, but also to program costs (e.g., the dollar gains in increased earnings attributable to a vocational rehabilitation program, as compared to the dollar costs of the program), benefit-cost analysis resulting in a difference between benefits and costs* (expressible in a single number) is possible. Then the implicit allocation decision is between the program being evaluated and all other uses of federal dollars (including lower taxes).

---

*More traditional practice has been to calculate the ratio of benefits to costs, but recent thinking suggests that the difference provides more information concerning alternative uses for the program funds in question.

In fact, meaningful benefit-cost computations are seldom feasible, but allocation to the program being examined as against other uses of the federal dollar is the implicit basis for many evaluations. The effort is to examine what charge--if any--against the agency budget, the Department budget, or the Federal budget is warranted. (See C, below--"For Whom" is the evaluation being made?). Evaluations whose primary purpose is to justify a program to higher authorities (if the evaluations come out right) are of this type, since the decision to be made--by the Office of Management and Budget or a Congressional committee, or whomever the program is being justified to--is a decision about allocation to the program as against a range of other programs.

b. Evaluation to Improve Strategies. This is the kind of evaluation designed not to see how well a program is doing as a whole, but to determine changes (alternative strategies) that will make it perform better. Such evaluations provide data for making decisions about program development and program revision. They may be designed to provide a basis for choice among techniques to be used within a project or program; or they may be designed to support technique refinement in an iterative process of evaluation and marginal change. Technique choice evaluations make a summary judgment about the relative effectiveness of new or existing alternatives, in a way similar to the allocation evaluations of whole programs discussed in (a). Technique refinement evaluations give the decisionmaker more or less continual feedback on the effectiveness of particular aspects of a project or program when it is in the formative stage.

Categories (a) and (b) are not mutually exclusive; a particular evaluation project may provide information of both types. Measurement of the effectiveness of Head Start in improving the learning capabilities of pre-school children, for example, may be designed to affect the allocation of anti-poverty funds to Head Start. However, if it compares Head Start effects on pre-schoolers of different ages in order to estimate the best age for initial intervention, it is also a strategic improvement evaluation dealing with technique choice.

In deciding whether an evaluation project or part of an evaluation project should be classified as an "allocation" or an "improvement" evaluation, the decision level for which the information is intended must be taken into account. For instance, an evaluation project might be designed to help project managers determine what makes the best projects so good; if so, it would be an improvement evaluation. Or it might be designed to help program managers determine which projects should be cancelled; if so, it would be an allocation evaluation. While in a brief description the two evaluations might sound very similar, they would in fact be quite different if appropriately designed to meet the differing decision needs. The one designed for improvement would concentrate on the mix of techniques within the projects and on environmental conditions confronting good and bad projects. The object would be to find out what techniques work best under what conditions, with a view of spreading the use of the best techniques under the appropriate conditions. The one designed to determine relative worth would concentrate primarily on the worst projects, and would attempt to compare the costs of cancellation with the costs of continued operation. It will frequently be economical, however, to look carefully at both sets of factors in order to provide inputs for both types of decisions.

2. Compliance Control

These activities, often called monitoring, are distinguished from evaluation proper by their attention to input measures (e.g., adherence to guidelines, workloads, administrative practices.) Such measures are different from the estimates of worth which are central to any definition of evaluation. The term "compliance control" is more descriptive of this category than "monitoring"; it is possible to monitor performance, but the activities normally referred to here monitor compliance --compliance with legislative intent and administrative regulations --not performance.

Under compliance control, two kinds of activities deserve separate mention:

a. Site visits by inspectors. (Not all site visits are for compliance control, however. Many evaluations, fitting within the strictest definition of the term, will include subjective components--estimates based on the close observations of trained evaluators.)

b. Routine reporting of input and workload data.

3. Capability Building

An evaluation office may engage in a number of activities designed to facilitate or improve the conduct of evaluation studies in the future. We refer to this sort of effort as "capability building." Three subcategories can be listed:

a. Developing a Data Base. Some data are useful for future evaluation activities, other data are not. While there will always be questions as to whether an investment in a data system should be listed under (2b) or (3a), it is important that relevant issues about such an investment be raised. One such issue is the distinction between collecting data for administrative use and collecting data for use in future evaluations. Another major issue is whether data systems should be set up now for evaluations to be carried out in the future, or whether ad hoc data collection efforts within future evaluations would be more efficient and economical.

b. Improving Evaluation Strategies and Methodologies. This category includes all activities directed at improving the state of the art.

c. Improving Non-federal Capabilities. This category includes all activities directed at disseminating the state of the art of other levels of government in order to help state and local decisionmakers --and private groups as well--to increase their abilities to carry out their own evaluations. One of the most important roles of HEW is as a service organization to public and private entities outside of the federal government. These groups have a major role in evaluating federal programs as they apply locally, and a major need to evaluate

their own programs, and HEW assistance can aid them in carrying out effective evaluation.

B. WHAT IS BEING EVALUATED?

The types of activities that might be evaluated can be classified as follows:

1. Major separable program(s); e.g., Title I of the Elementary and Secondary Education Act, Maternal and Child Health Care, Public Assistance.

2. Major component(s) of such a program; e.g., urban school assistance, prenatal health care, Aid to Families with Dependent Children (Unemployed Parents).

3. Operating project(s) within one of these programs; e.g., Title I assistance in Chicago.

4. Pilot or demonstration project(s); e.g., Demonstration Neighborhood Health Centers. (The evaluation of pilot or demonstration projects which is frequently within the mandate of an evaluation office, is included here: the setting up of such projects is not included within this listing. Evaluation of a pilot or demonstration may be used for purposes of allocation--should the program being tried on a pilot basis be given the resources to become a large-scale effort?--or for improvement--should the technique being tried be replicated throughout an existing program?)

C. FOR WHOM?

Different kinds of evaluations are useful at different levels in the decision structure, as the example on the relative effectiveness of projects illustrates.

For most purposes, the relevant decision levels are, in descending order:

1. The Secretary of Health, Education and Welfare

2. The Surgeon-General, the Commissioner of Education, the Administrator of the Social and Rehabilitation Service, the Commissioner of Social Security

3. Bureau Chiefs

4. Program Managers

5. Project Managers

6. Non-federal Decisionmakers. As noted above, public and private groups outside of the federal government have a need for evaluations of their own programs. In many cases, the most efficient way to carry out such evaluations is for federal evaluators to execute them. Particularly in the case of improvement evaluations, HEW may want to evaluate a new technique, which, if the evaluation is positive, will be useful across a broad range of programs and localities.

III. ANNUAL EVALUATION PLANNING

An evaluation plan lays out in advance the activities that an office responsible for evaluations expects to carry out during the period for which the planning is being done. In preparing such a plan, the responsible office can allocate the resources available for evaluation--both dollars and skilled manpower--in a way that puts first things first and that does not make commitments to relatively low priority evaluation efforts before funds are set aside for the most important ones. In addition, the evaluation plan provides those to whom the evaluation office is responsible with an opportunity to review projected evaluation efforts in the light of broader sets of needs. An evaluation plan is not a straitjacket, it is a guide. It should be considered changeable--because priorities have changed, because possibilities have changed, because new evaluation needs have been identified. But the plan is a discipline, and an aid to self-discipline. When new needs do arise, consideration will have to be given to whether they can be filled by obtaining additional resources for evaluation, or whether previously planned lower priority items will have to be dropped or postponed.

The planning formats and tabular summaries laid out here are designed to be useful to decisionmakers and to evaluators, to be simple and clear, and to minimize the duplication of planning efforts. They minimize the back and forth flow of paper by confining routine flows to summaries and brief abstracts. These short documents can then be used to flag the availability of more comprehensive statements retained in the files of the preparing offices and available upon request to responsible administrators and higher evaluation echelons.

To these ends, what is suggested is an annual evaluation plan, timed so that the preparation and review of the plan will be completed before it is necessary to make commitments to the planned projects. Although evaluation and evaluation planning must be matched to other time-cycles as well--the budgeting cycles of the programs being evaluated, for example--this can be accomplished by preparing different

simple summaries of the single plan, rather than additional plans. The summaries will also facilitate after-the-event review of the evaluation plan. At the end of each year, the summaries of the evaluations planned can be compared with summaries of evaluations actually carried out, to discover how effective evaluation planning actually was.

To achieve this, then, <u>each office responsible for evaluation should prepare an annual plan consisting of three parts</u>:

- o A statement of the strategic objectives of the evaluation plan.
- o A brief on each planned evaluation, to be prepared <u>only</u> for those evaluations for which the office has primary responsibility.
- o A number of aggregated tabular summaries, covering both the evaluations for which the office has primary responsibility, and those for which it has overall review responsibility.

A. STATEMENT OF STRATEGIC OBJECTIVES

The statement of strategic objectives should specify the programs and program areas for which the plan-preparing office has the <u>primary</u> responsibility for initiating, contracting, and supervising evaluations; and those for which it has the <u>review</u> responsibility for overseeing evaluations developed by offices in subordinate agencies.

For example, the Assistant Secretary for Planning and Evaluation has the primary responsibility for developing evaluations cutting across more than one HEW area (e.g., education and vocational rehabilitation), and also for carrying out specified evaluations for the Secretary. In addition, he has review responsibility for evaluations developed by the evaluation offices of the Office of Education, the Health Services and Mental Health Administration, and so on.

The statement should then lay out in the light of these responsibilities, the evaluation strategy for the year being planned. It

should list by rough priority ranking the programs and program areas to be stressed in the year's evaluation; it should give the reasons for the ranking in the light of the program decisions that must or should be made; and it should explain how the decisions may be affected by the evaluations. In addition to specific program evaluations, the statement should discuss and put into the ranking any compliance-control projects (A2 in the typology) for which the office is responsible, and any capability-building efforts (A3)--important for long-run program and evaluation purposes--that should be initiated during the planning year. The statement should discuss in general terms the kinds of evaluations that will be carried out to meet the objectives, their scope (e.g., single program or cutting across more than one program), their method, and their timing, and should make clear for which ones the office preparing the plan has primary or review responsibility.

B. EVALUATION BRIEFS

The plan should include a separate brief on each evaluation for which the office preparing the plan expects to take primary responsibility. A brief prepared as part of the plan should ordinarily list:

1. The program or program area being evaluated.
2. The evaluation category in terms of the typology of Section II, above.
3. A short description of relevant policy decisions and of the ways in which different possible outcomes of the evaluation may affect the decisions.
4. The timing of the evaluation: when it is expected to be initiated and on what schedule the information it is designed to produce is likely to become available.
5. An estimate of the number of dollars expected to be obligated in the initial year and in subsequent years.
6. An estimate of the number of man-months of evaluation office supervision required in the initial and in subsequent years; and, if possible, the name of the initial supervisor.

In addition to these briefs--which should in fact be brief--the specific evaluation rationale suggested in Section IV of this Guide for use in preparing any single evaluation, and the after-evaluation analysis also suggested there should be filed in the preparing office with the relevant brief, as the rationale and analysis become available. These rationales and analyses should, of course, be available to administrators and other evaluation offices in the supervisory chain; but to cut down on the flow of unnecessary paper, they should not be forwarded routinely. Any arrangements made on review of amendments to the evaluation plan should be decided by the relevant offices.

C. TABULAR SUMMARIES

As noted, the individual evaluation briefs should be included in the plans only of the offices with primary responsibility for these evaluations. Frequently, however, an evaluation office not only has direct responsibilities, but also review responsibilities over the evaluations carried out by a number of other offices. The simplest way to fulfill the planning requirements inherent in such overall responsibilities is to tabulate all the evaluations for which the office puttting out the plan has primary or review responsibility.

1. Basic Format

Sample Table 1 provides a basic format for such tabulations. By listing a year's evaluations vertically by program or component being evaluated, and horizontally by purpose of evaluation, and summing both ways, it provides an easy but detailed picture of allocation of effort by major categories. The completion date and the decision level for which the evaluation is being conducted are also indicated in this table.

An individual evaluation project may fit into more than one category; e.g., an evaluation may (and frequently should) be designed both to affect resource allocation, and to improve program strategies. In these cases, the dollars allocated to a particular project should be apportioned across two (or more) relevant categories. Hypothetical project 867 provides an example in Table 1.

Sample Table 1

(EVALUATION LISTING)

Planning Year (e.g. 1972)

Same Form Last Complete Year (1970) and Current Year (1971)

WHAT IS BEING EVALUATED ?	PROJECT ID. NO.	COMPLETION DATE	EVALUATION				COMPLIANCE CONTROL				CAPABILITY BUILDING				Total by Level Within Program
			ALLOCATION		IMPROVEMENT		SITE VISITS		REPORTING SYSTEMS		DATA SYSTEMS		STATE OF ART	HELP to STATE & LOCAL	
			For Whom?	$ (1000's)	For Whom?	$ (1000's)	For Whom?	$ (1000's)	For Whom?	$ (1000's)	For Whom?	$ (1000's)	$ (1000's)	$ (1000's)	
PROGRAM A: EDUCATION PERSONNEL DEVELOPMENT															
1. Total Program:	645				Prog Mgr	60									60
2. Major Component(s): Meeting Teacher Shortage	867	3/73	Marland	50							Future evaluators	25			75
3. Operating Project(s):	341	9/72					Prog Mgr	25							25
4. Demonstration Project(s): Project Phoenix	921	1/74			Prog Mgr Proj Mgrs	40									40
Total by Purpose Within Program:				50		100		25				25			200
% of Program Evaluation Budget:				25%		50%		12.5%				12.5%			
PROGRAM B: ELEMENTARY & SECONDARY EDUCATION															
•															
•															
•															

Each annual evaluation plan should include separate tables of this type on:

a. Planned activity in the year for which the plan has been prepared (e.g., 1972, in Sample Table 1).

b. Best estimate of actual activity in the current year (during which the plan has been prepared) (e.g., 1971).

c. Actual activity in the last complete year (e.g., 1970).

d. The plan that had been laid out for the last complete year when it was the planning year (e.g., the plan prepared in 1969 for 1970). (This table of course cannot be included until two years after the system suggested here has been adopted. Its purpose is for comparison to the table suggested in (c), to examine how closely the evaluation plans are actually being carried out.)

2. Summary Tables

Sample Table 2 provides a format for a series of planning summaries of use both to the evaluation office and to various review levels. Without listing each separate project as was done in Table 1, Table 2 tabulates vertically the total evaluation effort by major program and by type of evaluation (or compliance control or capability building) (Breakdown A in the typology of Section II.) This can be done on a single sheet for a number of years, and can include both dollar and manpower allocations, useful for different sorts of allocation.

Tabulations in the Table 2 format can be adapted and used flexibly, summarizing in different ways for different purposes. Initially, the following three versions of Table 2 are suggested for inclusion in the plan:

a. A summary of the evaluations to be initiated during the planning year (and other years). For the planning year, the total of dollars obligated should equal the total available for obligation. (It is perfectly legitimate, of course, to include an unallocated contingency fund.)

Sample Table 2

(EVALUATION SUMMARY) FOUR-YEAR EFFORT

Purpose	Last Complete Evaluation Year (e.g., 1970)		Current Evaluation Year (e.g., 1971)		Planning Year (e.g., 1972)		Next Year (e.g., 1973)	
	$ (1000's)	Man-Mo. of Supervision	$ (1000's)	Man-Mo. of Supervision	$ (1000's)	Man-Mo. of Supervision	$ (1000's)	Man-Mo. of Supervision
Program A. Education Personnel Development								
1. Evaluation								
a. Allocation	0		50				10	
b. Improvement	50		100		150		75	
2. Compliance Control								
a. Site Visits			25		20		10	
b. Reporting System								
3. Capability Building								
a. Data Systems			25		10		5	
b. State of Art								
c. Help to State & Local								

•
•
•
•

b. A summary of the evaluations that will be on-going during the planning year and subsequent years. This is a cumulated level-of-activity table: for the planning year it should include both evaluations begun in previous years that will still be on-going, and evaluations expected to begin in the planning year. For subsequent years, it should carry the implications of current and past plans for evaluations to be on-going then. For the planning year, the total of man-months of supervision to be used should be no greater than the total staff-time available for this purpose, allowing again for contingencies.

c. A summary of the information that will become available during the planning year and subsequent years. In this summary, evaluations should be tabulated by the year in which the final report is expected to be available; if interim reports providing enough information to affect decisions are also expected, they should be tabulated also. Since reports may be listed more than once, column totals in this context are not likely to be very meaningful.

For each of these tables, the first year listed should be the year completed before the date of the plan; actual experience should be tabulated in that column. For the year during which the plan is being prepared, the information should be based on best estimates. How far beyond the planning year a table should be carried depends on the purpose of the table in question. A summary of evaluations to be initiated may go no further than the planning year, although if consideration has been given to subsequent years, additional columns may be useful. A summary of evaluations on-going should be carried out for as many years as any current evaluations are expected to continue; similarly with a summary of information to become available.

The purpose of tables like these, then, is twofold:

- To facilitate easy reference to important data about the course of evaluation and its expected effect on decisions.
- To allow evaluation offices with review responsibilities to oversee these responsibilities simply by constructing broad aggregate summaries from the narrower summaries provided them by the offices being reviewed, rather than by duplicating the entire planning process at every level.

IV. THE RATIONALE FOR A SPECIFIC EVALUATION

This section is aimed at assisting evaluation offices, first, to lay out in advance the objectives of a proposed evaluation, and, second, to assess in retrospect how well the evaluation met these objectives. Advance answers to questions A through L, below,* will provide a record of the rationale under which the evaluation was initiated: after-the-event answers, will record an assessment of the degree to which the expectations expressed by the initial answers were fulfilled. As suggested in the discussion of evaluation planning, the records provided by these answers should be available to other responsible offices, but they should not be passed around routinely. Together with the evaluation reports themselves, they provide the basic record of each evaluation.

For each evaluation project, then, the following information should be provided--for the record, before starting and after completing an evaluation--by the office initiating the evaluation. These records should prove valuable to the staff members setting up and monitoring the project, to those budgeting for evaluation, to those involved in review, and to those who ultimately will assess the evaluation effort itself.

A. WHERE IN THE TYPOLOGY OF PART II DOES THE EVALUATION FIT?

For what decision-related purpose is the evaluation being carried out? What is being evaluated and for whom? It should be possible to code any evaluation, using the codes drawn from the typology, according to the answers to these questions. Such codes never fit every case completely comfortably (and, as has been noted, use will undoubtedly indicate needed changes in the typology). For this reason, the major information in the rationale documentation will be provided by verbal answers to the remaining questions. Nonetheless, codes are needed for purposes of summation in the planning and other processes.

---

*The discussion outlines laid out here under headings A through L need not be followed in detail. They are intended only as guides.

B. WHAT DECISION OR DECISIONS MAY BE AFFECTED BY THE EVALUATION REPORT?

As has been emphasized throughout, the purpose of evaluating a government program is to affect program or policy decisions; unless an evaluation project is designed for a particular decision, the chances are that the evaluation results will not change anything or even be seriously considered by a policymaker. Typically, the decision will be one of the following:

1. Evaluation to Affect Allocation

a. Go/No-go. A major program will very rarely be terminated because of an adverse evaluation report. Once under way, programs develop a momentum of their own. The operators become entrenched, and the beneficiaries become an interest group. Under these circumstances, the most likely effect of adverse evaluation results on a major program is the discontinuation of a project or of a discrete activity within the program. When a no-go decision follows in the wake of an evaluation, it is apt to be because the results serve as merely a last (but perhaps crucial) piece of derogatory evidence. Similarly, decisions to undertake major programs are rarely based solely--or even largely--on favorable results of pilot project evaluation. Rather, such results may be a contributory factor in a decision or supporting evidence for a decision that has, in effect, already been made.

b. Other Increase or Decrease of Resources. Although an evaluation-based decision to terminate a major program is unlikely, if a program receives an unfavorable evaluation, its slice of the next budget is likely to be smaller than it would otherwise have been (although not necessarily smaller than it had been previously, if the program is on a growth curve). Programs that receive favorable evaluations stand to gain. Too frequently, however, evaluations that measure overall program impact are designed to influence go/no-go decisions, providing little clear guidance on where to cut back an inefficient program or how to expand a seemingly effective one. Thus, if an evaluation is designed for use in a resource allocation decision, the prospective results should tell more than whether the impact (or the cost of achieving it)

is great or small. They should provide information on what characteristics of the program cause the measured impact or lack of impact.

2. Evaluation To Improve Strategy

a. Technique Choice. Program operators are likely to be relatively receptive to evaluations that examine alternative techniques applicable within their programs. However, if such evaluations have allocation implications, as evaluation of educational voucher programs does, for example, because of the implied possibility of allocating funds away from public school systems, they are as politically and bureaucratically sensitive as any other evaluations.

b. Technique Refinement. Evaluation can be useful in helping to shape programs by providing feedback to program managers on how techniques and processes can be improved. Results provided iteratively can facilitate the evolution of an effective program design. But the search for such marginal adjustments should not be used as an excuse to avoid hard decisions; the evaluator must take care that, in supporting sequential minor decisions, he not become an accomplice to avoiding major ones.

3. Compliance Control

If the project in question is not an evaluation but a compliance control effort in fulfillment of part of the mission of the evaluation office, that fact should be noted and the requirement for the project should be stated. The remainder of Section IV of this Guide then becomes irrelevant.

4. Capability Building

If the project is designed to facilitate or improve the conduct of future evaluation studies by building a data base, by improving evaluation strategies and methodologies or by improving non-federal capabilities, the particular improvement to be achieved should be described. The question will be, what are the gaps, in data, in method, or in training, to be filled, and what are the needs of the evaluation office or other such offices--federal or other--for such capabilities?

C. HOW MIGHT POSSIBLE ALTERNATIVE RESULTS OF THE EVALUATION AFFECT THE DECISION?

1. Evaluations

For any evaluation, no matter what the type, the evaluation office should know in advance what kind of information is being sought--what program aspects are being measured. Ordinarily, it will be possible to estimate the range of possible results. The central question, then, is: How are different possible results likely to affect the decisions discussed in Subsection B, above? Is a finding of "No impact" likely to be a factor in an allocation decision; how large an impact is likely to be of significant importance in the decision? Are certain kinds of results likely to be rejected out of hand, or are others likely to be misinterpreted? What is needed here is a contingency plan for the evaluation itself, laying out the decision implications of the range of possible results.

2. Capability-building Projects

How are the results of a project designed to increase capabilities likely to achieve this end? In what kinds of future evaluation projects are the new data likely to be used, and how? After these questions have been answered, the remaining questions in Section IV of the Guide will not be relevant.

D. WHO WILL MAKE THE FINAL DECISION?

Often "final authority" is not a very useful concept in planning evaluations. In the strictest sense it is lodged too high--at the Presidential, agency head or Congressional committee chairman's level. For purposes of deciding whether and how to undertake a particular evaluation, the decisionmakers of interest are those who have the authority to make or are likely to recommend a decision based substantially on evaluation results.

1. Which Decisionmaker

a. In the Department of Health, Education and Welfare. The responsibility within the government for considering the substantive merits of an issue--as opposed to signing a directive--is usually not precisely placed. Although most decision documents require the signature of a high-level executive, some--like refunding of a small demonstration program--are actually decided much further down. Others--like determination of major program allocations--are decided in fact as well as in form at the executive level. Occasionally, it is impossible to determine in advance where the decision will be taken. Since effect on decision-making is central to mounting an evaluation, it is important to delineate the general focus of the decision as closely as possible, and to judge what factors are likely to determine who will finally make the decision. Will the nature of the evaluation results matter? What if the individuals involved change?

b. Elsewhere in the Federal Government. As discussed below under (I2), evaluations may be designed to assist Department officials to defend programs elsewhere--before the Office of Management and Budget or the Congress, for example. This is a perfectly normal and legitimate use of evaluation. It does not imply that the evaluation should be set up to produce biased answers; rather that a hypothesis of "favorable" results, if confirmed, can be useful at various levels of government. Deliberately biased evaluation, more often than not, has proven politically counterproductive. If a program is controversial enough to need such evaluation results, it generally has suspicious opponents who can throw the results into question (as indeed, they frequently do with the results of a completely objective evaluation). Therefore, the use of evaluation for program defense has more limitations than is sometimes recognized.

c. Outside the Federal Government. For programs run locally, in all three areas of health, education and welfare, many decisions are taken by non-federal or non-public officials. Particularly in the case of evaluations designed to select or improve techniques, results may affect policy at these levels. If the techniques are applicable on a national or regional basis, the Federal government may be the appropriate

level for initiating evaluations whose results will affect decisions taken at lower levels. The evaluation of a particular form of sampling check of welfare eligibility, for example, may provide information of importance to welfare authorities in thousands of counties. If decisions are to be made at levels like this, the key to evaluation effectiveness lies in the ability to disseminate results widely to the many decisionmakers affected. (See F5, below.)

2. Constraints on Decision

In assessing the likelihood that evaluation results will affect a decision, it is essential to consider what and who constrains a decisionmaker's latitude for action. The intention here is not to suggest that every evaluation must fit precisely within these constraints. Such constraints are neither unchangeable nor permanent, and one function of evaluation is to change the constraints. But failure to recognize the constraints that exist, and to understand how they apply to the particular evaluation in question, can cause the evaluation to become irrelevant to any real policy decisions. Since at different levels decisionmakers are affected in different ways, an enumeration of constraints must be specific to be useful.

a. Presidency. What is the President's position, if any, on the issue? Is he hemmed in by past actions or statements? Are relevant White House Staff committed one way or another? What about OMB?

b. Congress. Are influential members of Congress committed in a particular way for or against the program? How much have they influenced past program decisions? How have they received past evaluation results? Are there important local district implications?

c. Partisan Politics. Does either party have a proprietary interest in the program? Are there any major ideological issues? Is there strong organized bipartisan support or opposition?

d. Administrative Politics. Is a decision likely to impinge on the private preserve of another major agency? On a strong element within your own or another agency? Is there at present a clearly

established lineup of advocates and adversaries within the bureaucracy? If so, how rigid is it?

e. Federal/State/Local Power Balance. Is an evaluation decision likely to affect the balance? If so, how delicate is the balance? How important is it to the other fish the decisionmaker has to fry?

f. Pressure Groups. Are any professional, beneficiary, or other groups effective as program advocates or adversaries? If so, how do they derive their power? How will they react to evaluation?

3. Other Information Channels

As a rule, decisionmakers do not rely on a single source of information regarding program effectiveness. Information can be of high quality or low, can become available regularly or randomly, can range from highly subjective assessment to unprocessed data. Whatever the form, the existence and impact of other information channels should be taken into account when evaluation decisions are made. In particular, how will other information affect the credibility of evaluation findings? Is the information source reasonably objective? Is the source an advocate or an adversary? Is the other source unimpeachable in the decisionmaker's eyes?

E. HOW WILL THE AVAILABILITY OF EVALUATION RESULTS MATCH THE DECISION TIMETABLE?

1. When Will Decisions Be Made?

Obviously, the decision timetables are important to anyone who is responsible for evaluation. If evaluation results are to matter, they must be in the hands of the decisionmaker when he is deciding. This does not necessarily mean "the sooner, the better"--that the evaluator should attempt to have results available for the earliest significant decision. His first step should be to lay out the probable decision stream for the program to be evaluated. There will be differences in decision types (B, above) and levels (D), as well as differences in timing.

a. Major Decision Points. By definition there are few of these in the life of a program. They include such events as program initiation, program termination, major expansion or contraction, expiration of authorizing legislation, and Executive Branch reorganization. Their timing may, in turn, be related to other, more regular events, such as the State of the Union address, Presidential Messages, and the end of a fiscal year. Decisions made at these points can be affected by evaluation results as well as by many other factors. The evaluator needs to know when the major decision points fall. What might affect their timetable? What is the likelihood that the decision will be not to decide?

b. Regular Periodic Decisions. Most of these decisions relate to the budget cycle. The President submits his budget to the Congress at a time specified by statute. Agency submissions to OMB must conform to an annual budget calendar. Decisions to seek funds for a new program or to reallocate the agency's budget among programs are also made annually, usually after the agency's appropriation bill has been passed and signed by the President. Finally, a cleanup of funding decisions at the project level (including evaluation projects) occurs during the month of June, so that agencies will not lose funds that must be obligated before the end of a fiscal year. Which of these periodic decisions might be affected by evaluation results? How predictable is the occurrence of major decisions in the cycle? How far in advance?

c. Incremental Decisionmaking. It is often difficult--even in retrospect--to establish the time at which important decisions on a program are made. Programs often grow along a path of very small decisions. Obviously, the technique-improvement class of evaluations is tailored to this pattern of decisionmaking. However, in some cases an evaluation may be undertaken precisely for the purpose of disrupting the incremental pattern. The Westinghouse Head Start evaluation, for example, was undertaken primarily in order to disrupt a pattern which tried to improve Head Start, without ever questioning its basic favorable effect on children or the substantial fund allocation based on this presumed effect. The faults of this evaluation in failing to suggest improvement strategies, as well as its partial success in disrupting

the allocation assumptions, were largely caused by the evaluator's desire to disrupt. To be effective, however, such disruption requires a good deal of skill on the part of the evaluation office and a thorough understanding of the decision environment. Why does the particular pattern of decisionmaking exist? Does evolution appear to have led to a satisfactory rate of improvement? What are the likely costs of giving up incrementalism? What is the most significant decision that can be made without disturbing the incremental pattern?

2. How Will the Evaluation Results Match This Decision Timetable?

The timetable for results will be determined to some degree by the way in which the evaluation is designed and the contract written. Hence, the evaluation office has considerable latitude for influence. For some cases, the evaluator should follow "the sooner, the better" principle. For others, the cardinal principle will be "if you want it bad, you get it bad." It is hard to formulate a general rule for setting up a timetable of evaluation results. The one worst decision on timing, however, is no decision--neither an early evaluation nor the necessary preparation for a later one. Because of the incremental decisionmaking pattern of Head Start referred to above, no allocation evaluation had been mounted early in the program and no early plans laid for a later one, so that when a clear need for such an evaluation was manifest, it had to be done in a relatively quick and dirty way.

a. Interim Findings. It is almost always possible to obtain some results from an evaluation effort before the work is complete. It is not always wise to do so. The evaluation office should prudently assume that interim evaluation results (as well as final ones) will leak out to advocates and adversaries alike. In the press and in the Congress, qualifying terms like "provisional," "incomplete," "raw," and "not yet analyzed" tend to be dropped when interim results are cited. Those who stand to be damaged by the evaluation may have some success in discrediting the whole evaluation on the basis of interim results. Usually, decisions are made to proceed with social programs (such as the Family Assistance Plan) before they are fully field-tested. In such cases, even interim results can be useful in program design.

The evaluator needs to judge the probable soundness of interim results. How could they be used to good purpose and bad? How well can they be protected? Will they derogate from the final results?

b. Final Results. The final results of most major evaluation projects should be targeted to a specific decision timetable. As a general rule, the sooner final results are available and can fit into the decision timetable, the better, and the near-term perspective of most government executives will be a force in this direction. There are some important exceptions to the rule, however. In some cases, longitudinal studies spanning several years may be preferable to studies whose final results will be available sooner. This is likely to be the case for a new program that appears sound but has many detractors on other grounds. In addition, during the first year or so, any program encounters start-up problems, and the results of overall impact evaluations conducted early are unlikely to reflect the program's realizable potential. Under such circumstances, any evaluation efforts offering results in the formative year should focus on technique choice or improvement, with overall impact best examined by means of longitudinal studies (which should be planned and begun early so that they can provide timely results for allocation decisions later on, however). The problem is to determine how to fit the final results into the decision timetable for the most impact. What wrong decisions might be made if final results slip in time? If they become available prematurely?

c. Salvage Value. Predictions of decision timetables cannot be made with great certainty--even when there appears to be a regular cycle. Hence, it is worthwhile considering the "salvage value" of an evaluation. What if its results are not available when the targeted decisions are made? What if the particular issue it addresses never arises? What will the evaluation results reveal about the broad social problem area and about useful (or fruitless) ways of attacking it?

F. WILL THE EVALUATION RESULTS BE READ AND UNDERSTOOD?

Obviously, if evaluations are to affect decisions, someone with a major role in the decision process must read and understand them.

This does not necessarily mean reading and understanding an entire report. The higher the level of the decisionmaker, the less time he is likely to spend on it. Yet, if he is to use the results to defend a decision against critics, he will have to spend enough time to make his defense effective. It is the responsibility of the evaluation office to supply material that is intelligible to the decisionmakers --the decisionmakers cannot be called upon to interpret esoteric or irrelevant material. The question, then, is, will the report be read by:

1. The Final Authority?

Executives at this level rarely have the time to read and understand evaluation reports--or even summaries. However, if a high-level executive is to be publicly credited with deciding on the basis of "objective" evaluation results, he should understand the conceptual basis for the evaluation and the general findings. This is likely to take at least an oral briefing.

2. By Those Who, in Effect, Decide?

Usually more than one individual may have authority to decide, and several may participate (see D, above). The evaluation report should be designed with all these potential decisionmakers in mind. They should be able not only to read and understand it, but also to defend the method and results against all but very technical criticism --e.g., in a Congressional hearing or before a news conference with press specialists.

3. By Those Who Advise and Inform the Decisionmaker?

Normally, the evaluation results will be brought to a decisionmaker's attention by his own staff or other trusted advisers. The chief of the evaluation office is likely to be among this group, or at least to be influential with this group. (If not, the chances are that evaluations won't matter much.) At least one member of this group should understand the evaluation study well enough to defend it against a technical assault and keep the boss from using the results incorrectly.

4. By External Individuals or Organizations?

It is prudent to assume that any evaluation done by a contract evaluator (and many done internally) will be read by Congressmen or Congressional staff members who have an interest in the program. If it is of sufficient importance, the evaluation will attract the interest of the press. A Congressional committee member or staffer is likely to be interested in subjects under his jurisdiction, a Congressman or reporter in projects within his area. In recognition of the iron law of the leak (as well as the Moss Freedom of Information Act), many agencies routinely make contractor evaluations public. But if the evaluations are important, they will be made public, whether routinely or not. They will thus be read externally; whether they will be understood is another question.

5. By Anyone Who Should Use the Results for Program Purposes?

Many federally funded social programs are operated outside the federal structure--by state or local governments, or under private contract. If the evaluation is supposed to affect decisions in such remotely run programs, as many are, a mechanism to disseminate the results is crucial to the effectiveness of the evaluation and to HEW's important service role. The evaluation rationale should specify in some detail the means of dissemination of results.

G. ARE THE EVALUATION RESULTS LIKELY TO BE MISINTERPRETED?

The concern here should be whether or not reasonable men are likely to interpret the results differently or whether they can easily misinterpret results for their own special purposes. (See F2a above.) Issues of methodology are obviously involved, but the focus here is on the non-technical readers who are unlikely to delve very deeply into methodological issues or even to understand them. The evaluation may be methodologically sound, but the results may still be subject to misinterpretation or misuse by someone who does not read deeply into a report.

1. Proxy Values

Since it is rarely possible to obtain direct measures of program outputs--measures that coincide with stated program objectives--most evaluation results are proxies or indirect measures. The objective of anti-poverty programs, for example, has been to raise the standard of living of those at the lower end of the income scale. The proxy measure adopted for this, however, was the number of families crossing a purely arbitrary level of dollar income called the "poverty line." The problem, then, is that a proxy may carry one connotation for one person and quite a different one for another. The creation of the "poverty line," in the example, engendered endless debates over whether the arbitrary standard had been set at the "right" point. Because of this controversy, too little attention was focused on the number of people moving across the line as a general proxy indicator with which to evaluate how well anti-poverty programs were achieving their complex income-raising objectives. The evaluator's aim cannot be to remove ambiguity from proxies; that would be impossible. However, the evaluator should consider whether a proxy might cause an unfortunate "knee-jerk" reaction. How clear is it which objectives the proxies serve? Are any of the proxies commonly used to represent an objective outside the program? Are any of the proxies likely to kindle the prejudices of someone with a major role in the decision?

2. Partial Results

If the evaluation comprises a number of separate parts, the chances are that partial results will be considered. How damaging would this be? Would it lead to a wrong conclusion? Would the overall impact of the evaluation be lost?

3. Generalization

Choosing a sample in a way that will lend broad significance to evaluation findings is a major methodological problem. Apart from this, however, the evaluator should consider whether a decisionmaker who pays little attention to methodology will construe the results either too

broadly or too narrowly. Is there anything about the evaluation or the decision situation that suggests generalization beyond what is valid? For example, might enemies of a program seize upon an adverse evaluation of a project to discredit the whole program? Is an important program advocate likely to reject valid generalization--say, because the sample did not cover the cities with "good projects"?

H. IF AN EVALUATION-BASED DECISION IS MADE, CAN IT BE IMPLEMENTED?

If an evaluation shows that certain kinds of counseling techniques are effective in a wide range of vocational rehabilitation programs, for example, and if the responsible authorities in Washington decide to adopt these techniques, will this be implemented in the field? A simple directive may not do it; dissemination of results of the evaluation may convince some local program administrators, but not others. Perhaps it cannot be done at all--at least in the short run. In one sense, implementation is the decisionmaker's problem and not the evaluator's. However, so many program decisions have run aground on implementation that an evaluation design must pay some attention to the feasibility as well as the desirability of change.

I. WHY DO DECISIONMAKERS WANT THE EVALUATION RESULTS?

Decisionmakers do not always seek evaluations because they are in search of objective truth. Indeed, the state of the art of evaluation is not so high as to inspire unmitigated confidence in the results of even the best projects. Decisionmakers often have a policy goal in mind when they ask for an evaluation. The evaluation office should consider this in assessing the decision relevance of prospective evaluations. Most situations will fall within the limits described below:

1. "I want to Find Out What the Program Has Done"

Devising programs that will alleviate major social problems is an uncertain business. The reason for evaluations is to find out what does or does not work. Ordinarily, however, decisionmakers have preconceptions about answers to the questions addressed by an evaluation. From the standpoint of the evaluation office, the best potential

consumer of results is one who had not made up his mind; the second best is one whose views are subject to change. A decisionmaker with strong a priori views on all programs will be a good customer for evaluation only when it supports these views. Does the decisionmaker have a stake in the program or a strong view on it, as it now stands? Is he inclined to look for new ideas. How fertile is the field for acceptance of fresh approaches?

2. "I Want Some Ammunition with which to Defend/Attack the Program"

As has been suggested, many evaluations have at least the implicit purpose of providing a committed decisionmaker with ammunition to defend a program at a higher level--the Administrator before the Secretary, the Secretary before OMB or the President or the Congress. Conversely, some approaches to social problems are generally discredited. Yet they persist in programs, and evaluations of such programs may be undertaken with the sole objective of adding weight to the argument for abolishing them, this fact being only thinly disguised. Neither of these implicit objectives, of itself, destroys the objectivity of an evaluation, although it may impair its credibility. However, there is cause to be concerned about the decision relevance of an evaluation if, first, its major supporter is someone who is principally interested in the support it will give him and, second, the outcome of the evaluation is uncertain. How unpredictable are the results? Does your decisionmaker-sponsor understand this as well as you do? How will he react to the results if they do not support him? Is an end run feasible, if he tries to suppress them? Would it be worthwhile?

J. WHO HAS THE MANAGERIAL AND TECHNICAL COMPETENCE AND TIME TO DO THE EVALUATION?

Sound evaluations of social programs demand imaginative management and imaginative use of professional skills. These qualities are often hard to find in combination; where they are found together, the person, government office or contractor tends to be overcommitted. These factors as well as credibility (discussed below) should determine the choice, or division of effort, between government and outside evaluators.

1. In-Government Evaluation

a. Single Agency. Which of the agency's employees will do the work? Are they competent and sufficient in number? Does the team leader have the managerial and bureaucratic skill to deal with problems that may arise?

b. Interagency Task Force. Who will lead the task force, and how much time can he devote to the work? What kinds of representatives will the other agencies provide? How firm is the claim on their time? Will parochial interests of the agencies interfere with effective use of their representatives' professional skills? Caution: An agency is usually unwilling to provide a significant amount of a good man's time unless it wants to represent its particular interest.

c. Consulting Help. Will the government evaluation effort rely on significant help from outside consultants? Are consultants sure to be available when needed? Will they function well under the guidance of the evaluation leader, or might they simply "do their thing?"

2. Contract Evaluation

a. Competitive Bid. Are several contractors capable of designing an evaluation within carefully set terms of reference, of carrying it out, and of working with the in-house staff? Is the in-house staff knowledgeable enough about the program and evaluation methodology to set terms of reference which will provide a firm basis for submission of competitive bids?

b. Sole Source. Is only one contractor equipped to carry out work of the kind and quality desired by the contract office? Or is the project design so complex that no competent contracting organization will devote resources to producing a pre-contract evaluation design proposal that can be assessed on its merits unless assured that it will be the final contractor? Can such a sole source decision be defended, legally and politically?

c. If the evaluation is to be contracted out, is there sufficient in-house staff to design carefully the terms of reference, to monitor

the evaluation while it is in process, and to interpret the results when it is completed? This may be the single most important question in this guide. Evaluations do not carry themselves out, and even the best of contractors cannot substitute for in-house staff in the knowledge of the relevance of the evaluation to the decision process. Failure to provide adequate in-house control has almost always been at fault in the past, whenever evaluations have failed or turned out to be irrelevant.

K. HOW CREDIBLE IS EVALUATION?

Credibility has at least two dimensions--professional competence and freedom from bias. Both must be considered, whether the evaluation is done in-house or on contract. Professional competence is largely a question of reputation. If an organization--inside or outside of government--is identified with sound analysis, the issue of credibility usually will not arise on grounds of competence. (Of course, the finished product will have to bear professional scrutiny.) But the question of bias will almost always arise independently whether the work is done in-house or contracted out.

1. Credibility within the Executive Branch

Can the evaluating office and the evaluator be considered sufficiently unbiased? This is a complex question. The Director of the National Institutes of Health may trust his evaluation office to tell him what is good or bad, or what should be changed in health research programs; the evaluation chief reports to him, not to the administrators of the separate programs being evaluated. But if the Director uses such an evaluation at a higher level to back up a program for which he is responsible, the Secretary of HEW may not accept the report as free from bias. To the Secretary, the Director of NIH is necessarily an administrator evaluating his own programs--far from a disinterested party. Or if the Secretary does credit the Director's lack of bias, the Director of the Office of Management and Budget may not. This sort of problem goes up and down the line. It is asymmetrical, since it is

the favorable evaluations that are most likely to be blessed by program administrators and least likely to be trusted by outsiders.

2. Credibility on the Outside

When an evaluation reaches the Congress and the public, will it be considered unbiased? The answer is likely to be "No" for an evaluation provided by the Executive Branch, for the reasons discussed above. This fact of life should not deter the carrying out of evaluations. However, it does suggest that an evaluation designed primarily to convince the Congress of the viability of a program will seldom work.

L. HOW DOES THE COST OF THE EVALUATION COMPARE WITH THE PROGRAM DOLLARS TO BE AFFECTED BY THE DECISION?

If the two are even close, the evaluation should not be undertaken.* Estimating program dollars to be affected by the decision is difficult and enumerating evaluation cost is far from trivial, although it must be done for contract work. The problem, as a whole, is akin to that encountered by a speculative investor matching his resources against investment opportunities: The evaluation costs are analogous to the speculator's resources (his time, his limited ability to analyze thoroughly all possible opportunities, his capital); the program dollars are analogous to the prospective return on the investment.

1. Evaluation Costs

Contract dollars can be estimated initially or put in as a ceiling. They can be tied down firmly before a decision is made to go ahead. Costs in terms of evaluation office personnel resources are harder to estimate, because of the difficulty of predicting the bureaucratic and other problems an evaluation will run up against. It is usually unnecessary to attach a dollar figure to the office's personnel costs as

---

*An exception to this rule might be when the evaluation has some general applicability and is a way of getting started.

long as the total time constraint is considered.* However, where there is a tendency to do in-house evaluations "to save money," it may be useful to rack up office personnel costs against the total staff budget.

2. Program Dollars

Obviously this estimate is closely linked to types and levels of decisions to be effected and targeting of the evaluations to particular decision points. Like other "risk investment" problems, it is usually possible to identify the several most likely outcomes, and it is worthwhile to attach some subjective probability to each "return." For an evaluation bearing in on allocation decision, for example, the likely range of budget variation, given alternative evaluation outcomes, should be estimated. (For a demonstration project, the relevant budgets are those that might fund a general adoption of the program being demonstrated, if the evaluation shows effectiveness. The cost of the evaluation may thus be large relative to the overall project cost.) In a technique-improvement evaluation, estimation should be relatively easy if the technique is designed to be cost-saving. If it is an effectiveness-increasing technique, it may be useful to estimate the size of the program to which an improved technique will apply, although a direct comparison to costs will not be very meaningful.

---

*Evaluation experience in the Office of Economic Opportunity suggests a rule of thumb: a mid-level member of the evaluation staff (GS 12-14) should be assigned to monitor each $500,000-800,000 of annual evaluation work (two to five large projects).

V. THE CRITICAL PROGRAM SUMMARY

Although the suggestions above for planning, for selecting, and for carrying out evaluations are centered on the role of the evaluation in the program decision process, they are nonetheless designed to structure the evaluation process itself. This section is concerned with programs, and uses formal evaluation as one instrument among several in judging the programs.

The Critical Program Summary is a matter of art, not of science. It depends heavily on judgment, the final judgment being that of the responsible head of the cognizant evaluation office. The Summary makes use of evaluation, but brings together data from many sources, not excluding personal beliefs, identified as such. The reason for the preparation of such summaries by evaluation officers is that the very process of planning, selecting, supervising, and carrying out evaluations gives these officers unique viewpoints that can be of substantial assistance to decisionmakers deciding on program actions.

The Critical Program Summary, then, is a document prepared by each evaluation officer for each program for which he is responsible, and submitted to both the administrative office to which the evaluation office reports and to the next higher evaluation echelon. The Summary should answer the following questions:

A. WHAT IS THE PROGRAM IN QUESTION, AND TO WHAT PROBLEMS IS IT ADDRESSED?

What is the problem? How many people does it affect (the "Universe of Need")? How does the program attempt to affect the problem? What portion of the Universe of Need does it cover? What is the legislative basis for the program, and the legislative history?

B. WHAT DO WE NOW KNOW ABOUT THE PROGRAM?

What is meant here is relatively hard knowledge. The question to be answered is, what knowledge has been provided us about the program from evaluations completed or partly completed? Such evaluations are,

of course, the major source of hard knowledge but other sources may be available too. For example, research reports discussing the problem to which the program is addressed may be available to the evaluation offices; non-governmental efforts such as journalistic reports, university research, and similar documents may be available. In any case, the answer laid out here should summarize everything that can be said about the program on the basis of relatively hard knowledge.

C. WHAT DO WE THINK ABOUT THE PROGRAM?

The answers to this question draw upon the hard knowledge just discussed, but this is one source among many, all drawn together by the informed judgment of the evaluation officer. In addition to hard knowledge, the evaluation officer may use quite imperfect data, anecdotal information, logical judgments about what seems to be the case, even defensible personal beliefs. What should be set forth here are judgments, then, but not unsupported judgments. They should be based on plausible arguments set forth explicitly, and they should summarize the evaluator to present his own informed beliefs about the past and current courses of the program in question.

D. WHAT ACTION DO WE RECOMMEND IN REGARD TO THE PROGRAM?

To answer this question, the evaluation officer must play the role of decisionmaker. Granted that the evaluator is not likely to bring to bear the full range of factors considered by the decisionmaker, he is nonetheless a public official whose judgments in response to questions (A) and (B), above, should logically lead to policy recommendations. The evaluator thus should make his recommendations explicit here: what programs and projects should be promoted or eliminated, what allocations should be made, what techniques should be promoted?

E. WHAT ADDITIONAL KNOWLEDGE WOULD WE LIKE TO HAVE ABOUT THE PROGRAM?

Questions (A) through (D) are policy questions, the answers to which will necessarily be based on partially informed judgment. In the very process of systematizing this judgment, however, gaps in knowledge

will be uncovered and spotlighted. In his answer to question (D), the evaluation officer has provided his best judgment as an informed official as to what should be done. The answer to the present question will identify further information that could cause him either to change his recommendation or to feel more comfortable in his initial judgment.

F. WHAT SHOULD BE DONE TO OBTAIN THIS INFORMATION?

Here the evaluator returns to his initial and primary role. Considering the information that he would like to have, what actions is he taking as an evaluator to obtain it and what actions might he take, or if he is not taking action, why not? What steps outside of the evaluation process proper--research, experimentation, demonstration--ought to be taken by others to obtain the requisite information?

The final document should be a bold one. Whereas evaluation itself must be relatively rigorous and eschew the highly conjectural, this document should be deliberately conjectural--the conjectures of a party with information and understanding that are different from and sometimes better than those of other parties to program decisions. The document that attempts to answer the questions of what should be done can be a major aid to those responsible for making these decisions --in some cases, perhaps, the greatest aid that can be provided by an evaluation office.

Appendix A

DOCUMENT SUMMARY

The various sections of this Guide suggest the following documentation:

1. An annual evaluation plan, prepared by each evaluation office, and containing:
   a. A statement of strategic objectives. (page 11)
   b. A brief on each planned evaluation. (page 12)
   c. Four basic tables:
      (1) Planned evaluation activity: planning year. (page 15)
      (2) Actual evaluation activity: current year. (page 15)
      (3) Actual evaluation activity: last complete year. (page 15)
      (4) Previously prepared plan: last complete year. (page 15)
   d. Three summary tables:
      (1) Evaluations initiated each year. (page 15)
      (2) Evaluations ongoing each year. (page 17)
      (3) Evaluation information becoming available each year. (page 17)

2. A statement of rationale of each evaluation, prepared by the initiating office two times. (pages 18-36)
   a. Before the evaluation is begun.
   b. After the evaluation is completed.

3. A critical program summary, prepared by each chief evaluation officer for each program for which he is responsible, prepared on a schedule dictated by the needs of program administrators. (pages 37-39)